**1. Dataset Information**

**Dataset Used**: *Friday-WorkingHours-Afternoon-DDos.pcap\_ISCX.csv*

* The dataset contains 79 columns, including 78 features and 1 target column, Label.
* The Label column identifies traffic as either **BENIGN** or an **attack**.

**2. Preprocessing Steps**

**Dataset Cleaning**

The raw dataset contained some issues:

1. **Null Values**: Certain rows contained missing values.
2. **Infinite/Extreme Values**: Features like Flow Bytes/s contained infinite (inf) or extremely large values.
3. **Redundant Columns**: All columns were retained since none were irrelevant.

**Cleaning Process:**

* **Remove Rows with Null Values**: Rows with missing values were dropped using:
* **Standardize Column Names**: Leading/trailing whitespace in column names was removed:
* **Replace Infinite/Extreme Values**: inf and -inf were replaced with NaN, and then these were filled with column means:
* The cleaned dataset was saved as **Cleaned Friday-WorkingHours-Afternoon-DDos.csv**.
* Cleaned dataset values have been standardized to avoid extreme values causing training issues.

**3. Analysis Workflow**

**Encoding the Target Variable**

The target column (Label) was encoded for binary classification:

* **BENIGN** → 0
* **Attack Labels** → 1

**Train-Test Split**

The dataset was split into:

* **70% Training Data**: Used to train the Random Forest model.
* **30% Test Data**: Used for evaluating the model’s performance.

**Model Training**

A **Random Forest Classifier** was trained using default parameters.

model = RandomForestClassifier(random\_state=42)

model.fit(X\_train, y\_train)

**4. Evaluation Metrics**

1. **Confusion Matrix**: A heatmap of the confusion matrix was generated to compare actual vs. predicted labels.
2. **Classification Report**: This report includes precision, recall, F1-score, and accuracy for each class.
3. **ROC Curve**: The Receiver Operating Characteristic curve shows the trade-off between true positive and false positive rates, along with the Area Under the Curve (AUC) score.

**5. Instructions for Running the Code**

**Prerequisites**

Install the required Python libraries:

pip install pandas numpy scikit-learn matplotlib seaborn

**Folder Structure**

Place the following files in the specified locations:

* **Dataset File**: Cleaned Friday-WorkingHours-Afternoon-DDos.csv
  + Path: C:/Users/devan/OneDrive/Desktop/RF/
* **Python Script**: Rf code.py
  + Path: Same folder as the dataset.

**Running the Script**

1. Open your Python IDE or terminal.
2. Execute the script

**Expected Outputs**

* **Confusion Matrix**: A heatmap showing the classification results.
* **Classification Report**: Precision, recall, and F1-score for each class.
* **ROC Curve**: AUC score and the ROC graph.